AD-A120 296 TEXAS UNIV AT AUSTIN F/G 12/1
CONVERGENCE OF VECTOR QUANTIZERS WITH APPLICATION TO THE DESIGN--ETC(U)
1981 G L WISE, E F ABAYA AFOSR-81-0047
UNCLASSIFIED AFOSR-TR-82-0821 NL

1 OF 1
AD A
120296

END
DATE
FILMED
11-82
DTIC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A120296

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 82-0821	2. GOVT ACCESSION NO.: AD-A120296
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): CONVERGENCE OF VECTOR QUANTIZERS WITH APPLICATION TO THE DESIGN OF OPTIMAL QUANTIZERS	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Gary L. Wise and Efren F. Abaya	8. CONTRACT OR GRANT NUMBER(s): AFOSR-81-0047
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Electrical Engineering, University of Texas, Austin TX 78712	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: PE61102F; 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Mathematical & Information Sciences Directorate, Air Force Office of Scientific Research, Bolling AFB DC 20332	12. REPORT DATE: 1981
	13. NUMBER OF PAGES: 10
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE OCT 14 1982 S D A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Presented at the Nineteenth Annual Allerton Conference on Communication, Control and Computing, 30 September - 2 October 1981, Monticello, Illinois. Published in the Proceedings of the Conference, pp. 79-88.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Quantization, convergence of a sequence of quantizers, empirical quantizers.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Suppose that a sequence of probability distribution functions $\{F_n\}$ converges weakly to a distribution function F. When does the sequence of optimal quantizers for the F_n's converge to an optimal quantizer for F? Sufficient conditions are given to guarantee this convergence for scalar and vector quantizers with a general class of distortion measures. These results are used to examine (CONTINUED)

DTIC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ITEM #20, CONTINUED: several aspects of quantization, including the existence of minimum r-th power distortion quantizers and the convergence of a recently proposed algorithm for designing optimal quantizers.

A

AFOSR-TR- 82-0821

CONVERGENCE OF VECTOR QUANTIZERS WITH APPLICATION TO THE DESIGN OF OPTIMAL QUANTIZERS

EFREN F. ABAYA
Department of Electrical Engineering
University of Texas at Austin
Austin, Texas 78712

and

GARY L. WISE
Departments of Electrical Engineering and Mathematics
University of Texas at Austin
Austin, Texas 78712

ABSTRACT

Suppose that a sequence of probability distribution functions $\{F_n\}$ converges weakly to a distribution function F. When does the sequence of optimal quantizers for the F_n's converge to an optimal quantizer for F? Sufficient conditions are given to guarantee this convergence for scalar and vector quantizers with a general class of distortion measures. These results are used to examine several aspects of quantization, including the existence of minimum r-th power distortion quantizers and the convergence of a recently proposed algorithm for designing optimal quantizers.

I. INTRODUCTION AND PRELIMINARIES

Suppose that a sequence $\{F_n\}$ of probability distribution functions on $\mathbb{R}^k$ converges weakly to a distribution function F. Under what conditions does the sequence of optimal block quantizers for the F_n's converge to an optimal quantizer for F? This question arises in the design of quantizers for unknown or incompletely specified distributions, where it is necessary to use either an estimate of the true distribution or an empirical distribution based on a set of observations. It may be desirable to know that if the estimated or empirical distributions converge to the true underlying distribution, then the quantizers produced approach optimality. The same question occurs in a recent paper on the design of vector quantizers [5]. In most cases, the algorithms for designing optimal quantizers that have appeared in the literature [1,5,6,8] are based upon some form of Max's conditions [8] which are necessary, but not sufficient, for optimality. As a result, the algorithms may converge to a quantizer which is locally optimal for the target distribution F, but is not globally optimal. If an algorithm were started sufficiently close to a global optimum, this problem would not arise. The following proposal by Linde, Buzo and Gray is intended to insure that this always happens [5]. Consider a sequence of distribution functions $\{F_n\}$ converging weakly to the desired distribution F, with F_1 having a unique locally optimal (and therefore also globally optimal) quantizer. If the F_n are taken to be sufficiently "close," then their respective optimal quantizers might reasonably be expected to be close to each other. So if Q_n, an optimal quantizer for F_n, is used as the starting quantizer for F_{n+1}, the algorithm is likely to converge to a globally optimal quantizer Q_{n+1} for

Presented at the Nineteenth Annual Allerton Conference on Communication, Control, and Computing, September 30-October 2, 1981; to be published in the Proceedings of the Conference.

F_{n+1}, despite the presence of other local minima. Intuitively, one might expect the sequence of quantizers $\{Q_n\}$ produced by this doubly iterative procedure to converge to a globally optimal quantizer Q for the limiting distribution F. In this paper, we will discuss conditions under which this convergence can take place.

First we give a few definitions and establish some notation. An N-level k-dimensional vector quantizer is a mapping $Q : \mathbb{R}^k \to \mathbb{R}^k$ which assigns to the input vector x an output vector Q(x) chosen from a finite set of N distinct vectors $\{y_i : y_i \in \mathbb{R}^k, i = 1,2,\ldots,N\}$. When optimal quantizers are being considered, there is no loss of generality in assuming the "nearest neighbor" assignment rule: Q(x) is that member of the set which is nearest to x in Euclidean norm, with ties being broken in some pre-assigned manner. This rule will be adopted throughout the rest of this paper. Scalar quantization (k=1) is considered to be a special case of vector quantization. Generally, the input X is a random vector taking values in $\mathbb{R}^k$ and having a probability distribution function F. The performance of a quantizer Q is measured by a probabilistic mean distortion measure

$$D = D(Q,F) = \int C_0(\|x - Q(x)\|)dF(x). \qquad (1)$$

Here $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^k$ and $C_0(t)$ is a non-negative cost function. For simplicity, we will sometimes write $C(x) = C_0(\|x\|)$. As a side result of this paper, we will show that the minimum of (1) can be achieved for the r-th power distortion $D = E\{\|X - Q(X)\|^r\}$. An optimal quantizer is one which minimizes (1) among the class of N-level k-dimensional quantizers.

Let F_n and F be k-dimensional probability distribution functions. The sequence $\{F_n\}$ is said to converge weakly to F (written $F_n \xrightarrow{w} F$) if $F_n(x) \to F(x)$ at every continuity point x of F. We say that $\{F_n\}$converges setwise to F ($F_n \xrightarrow{s} F$) if

$$\lim \int_B dF_n(x) = \int_B dF(x)$$

for every Borel subset B of $\mathbb{R}^k$. Setwise convergence of $\{F_n\}$ implies weak convergence. The following theorem (adapted from [9, p.232]) pertains to setwise convergence.

<u>Theorem 1.1.</u> Let $F_n \xrightarrow{s} F$, and let $\{f_n\}$ and $\{g_n\}$ be two sequences of real-valued Borel measurable functions which converge pointwise to f and g, respectively. Suppose that $|f_n| \le g_n$ and that $\lim \int g_n dF_n = \int g\, dF < \infty$. Then $\lim \int f_n dF_n = \int f\, dF$.

If $g_n : \mathbb{R}^k \to \mathbb{R}$ is a sequence of real-valued measurable functions, we say that $\{g_n\}$ is uniformly integrable with respect to the sequence of distribution functions $\{F_n\}$ if $\int |g_n|\, dF_n < \infty$ for all n and

$$\lim_{a\to\infty} \sup_n \int_{\|x\|>a} |g_n(x)|dF_n(x) = 0.$$

A "rectangle" $\overset{k}{\underset{i=1}{X}} [a_i, b_i]$ in k-dimensional space will be called a closed cell. On the real line (k=1) a closed cell is identical to a closed interval.

II. DEVELOPMENT

In this section k and N are fixed positive integers. Unless otherwise stated, X denotes a k-dimensional random vector; F_n and F are distribution functions on $\mathbb{R}^k$. Henceforth, it will be assumed that $C_0(t)$ is non-negative and nondecreasing on the ray $[0,\infty)$. Additional restrictions will be imposed as they are needed.

Under the nearest-neighbor assignment rule, a scalar quantizer Q can be represented by the N-tuple $(y_1, y_2, \ldots, y_N)$ of its output levels indexed in increasing order. With this representation, it is natural to say that a sequence of quantizers $\{Q_n\}$ converges to an N-level quantizer Q if the sequence of N-tuples associated with $\{Q_n\}$ converges to the N-tuple representing Q. This implies that $Q_n(x) \to Q(x)$ at all continuity points of Q. Motivated by this observation, we say that a sequence of vector quantizers $\{Q_n\}$ converges to a vector quantizer Q if $Q_n(x) \to Q(x)$ at all continuity points of Q. This defines the convergence of quantizer sequences. Notice that this definition allows the limit quantizer Q to have fewer than N levels. In any case, it follows from this definition that $C(x-Q_n(x)) \to C(x-Q(x))$ at every continuity point of $C(x-Q(x))$. In fact, if C_0 is continuous, then the above convergence is uniform on compact cells.

For the moment, assume that C_0 is continuous. Consider a sequence $\{F_n\}$ converging weakly to the distribution function F, in such a way that for each finite y in $\mathbb{R}^k$, $C(x-y)$ is uniformly integrable with respect to $\{F_n\}$. We will say more about this assumption in Section 3. Let $\{Q_n\}$be a sequence of N-level quantizers converging to the quantizer Q. Assume that Q has N output vectors $y_1, y_2, \ldots, y_N$. Imagine a hypercube (with sides of unit length) centered on y_i, and denote the midpoints of the 2k faces by z_i^j, j=1, 2,...,(2k). Then

$$C(x - Q_n(x)) \le \sum_{i=1}^{N} \sum_{j=1}^{2k} C(x - z_i^j) \tag{2}$$

for n sufficiently large. Therefore $\{C(x - Q_n(x))\}$ is uniformly integrable with respect to $\{F_n\}$. Applying Theorem A.1, given in the Appendix, we conclude that $D(Q_n, F_n) \to D(Q,F)$. In particular, if Q' is an N-level quantizer, then $D(Q', F_n) \to D(Q', F)$. Suppose that Q_n above is optimal for F_n. Then $D(Q_n, F_n) \le D(Q', F_n)$. It follows that $D(Q,F) \le D(Q',F)$. Thus the limit quantizer Q turns out to be optimal for the limit distribution F. As a matter of fact, Q need not have exactly N levels. If it had less than N levels, it would still be optimal for F among quantizers with N levels or less. In summary, we have the following result.

Theorem 2.1. Assume that $C_0(t)$ is non-negative, nondecreasing on $[0,\infty)$ and continuous. Suppose that $F_n \xrightarrow{w} F$, and that Q_n is an optimal N-point quantizer for F_n. If $C(x-y)$ is uniformly integrable with respect to $\{F_n\}$ for each y in $\mathbb{R}^k$, and if $\{Q_n(x)\}$ converges to some quantizer $Q(x)$ at all continuity points of Q, then Q is optimal for F.

The continuity requirement on $C_0(t)$ can be removed by strengthening the mode of convergence of $\{F_n\}$. For example, we can demand that $F_n \xrightarrow{s} F$. This implies pointwise convergence of $F_n(x)$ to $F(x)$, a stronger condition than weak convergence. However, if F and F_n have densities f and f_n, respectively, and $f_n(x) \to f(x)$ almost everywhere, then the two modes of convergence are equivalent. As usual, we assume that C_0 is non-negative and nondecreasing. Assume further that $C(x-y)$ is uniformly integrable with respect to $\{F_n\}$ for each y in $\mathbb{R}^k$, and that $\{Q_n\}$ converges to an N-level quantizer Q. For a fixed y,

$$\left|\int C(x-y)dF_n(x) - \int C(x-y)dF(x)\right| \le \left|\int_{\|x\|\le a} C(x-y)dF_n(x)\right.$$

$$\left.-\int_{\|x\|\le a} C(x-y)dF(x)\right| + \int_{\|x\|>a} C(x-y)dF_n(x) + \int_{\|x\|>a} C(x-y)dF(x).$$

Applying Theorem 2.1 to the first group of terms on the right-hand side and the uniform integrability requirement to the last two terms, we see that $\lim \int C(x-y)dF_n(x) = \int C(x-y)dF(x) < \infty$. Combining this with (2) gives a dominating function which satisfies the hypotheses of Theorem 1.1. Therefore $D(Q_n,F_n) \to D(Q,F)$. Proceeding as in the proof of the previous theorem, we arrive at the following.

Theorem 2.2. Assume that $C_0(t)$ is non-negative and nondecreasing on $[0,\infty)$. Suppose that $F_n \xrightarrow{s} F$, and that Q_n is an optimal N-level quantizer for F_n. If $C(x-y)$ is uniformly integrable with respect to $\{F_n\}$ for each y in $\mathbb{R}^k$, and if $\{Q_n(x)\}$ converges to some quantizer $Q(x)$ at all continuity points of Q, then Q is optimum for F.

It may be noted in passing that $C(x-y)$ is uniformly integrable with respect to $\{F_n\}$ if (a) the F_n's have uniformly bounded supports, or (b) $C_0(t)$ is bounded and $F_n \xrightarrow{w} F$. The key assumption of uniform integrability in the above theorems is not necessary, but cannot be completely dispensed with as the following example shows. Define

$$F_n(x) = \begin{cases} 0 & x < 0 \\ 1-1/n & 0 \le x < n \\ 1 & n \le x \end{cases}$$

$$\lim F_n(x) = F(x) = \begin{cases} 0 & x < 0 \\ 1 & x \ge 0 \end{cases} .$$

If $C(x) = |x|$ (mean absolute distortion), then $C(x-y)$ is not uniformly integrable with respect to $\{F_n\}$. Yet the optimum 1-level quantizer for F_n is its median, $\mathrm{med}(F_n) = 0$, which does converge to the median of F. On the other hand, if $C(x) = x^2$ (mean square distortion), the optimum 1-level quantizer for F_n is the expected value $\int x\, dF_n(x) = 1$. This does not converge to the expected value of F, which is $\int x\, dF = 0$. Of course, $(x-y)^2$ is not uniformly integrable with respect to $\{F_n\}$.

Common to the above theorems is the assumption that the sequence of optimal quantizers $\{Q_n\}$ converges. However, all that we really need is a subsequence of $\{Q_n\}$ which converges (to an optimal quantizer). This can be isolated in the following way. Consider the sequence $\{y_{1,n}\}$ of output levels with smallest magnitudes. If this has a finite limit point, then select a convergent subsequence $\{y_{1,n_i}\}$; otherwise, retain the original sequence. Now take the sequence $\{y_{2,n_i}\}$ of output levels with second smallest magnitudes and select a convergent subsequence $\{y_{2,n_{i_j}}\}$, if possible; otherwise, keep the sequence $\{y_{2,n_i}\}$. Do this for all the output levels in turn. At the end of this procedure, either a convergent subsequence $\{Q_{n_j}\}$ will have been selected, or it will have been determined that none of the sequences $\{y_{i,n}\}$ have finite limit points. We now show that this latter possibility can only arise in the trivial case that $C_0(t)$ is a constant with F-probability 1. Suppose that none of the $\{y_{i,n}\}$, $i=1,2,\ldots,N$ have finite limit points. Then

$$\lim_{n\to\infty} C(x-Q_n(x)) = \lim_{t\to\infty} C_0(t).$$

Call the limit on the right $C(\infty)$. If $C(\infty)$ is finite (i.e., C_0 is bounded) then it follows from the Dominated Convergence Theorem [9, p.229] that

$$\lim \int C(x-Q_n(x))\, dF_n(x) = C(\infty). \tag{3}$$

This implies that every quantizer with N levels or less has distortion $D = C(\infty)$. In particular we have $\int [C(\infty) - C(x)]\, dF(x) = 0$. Since the integrand is non-negative, $C(x) = C_0(\|x\|) = C(\infty)$, a constant, with F-probability 1 [9, p.228]. If $C(\infty) = \infty$ (i.e., C_0 is unbounded) it is just as easy to show that (3) holds, which implies that every quantizer with N levels or less has infinite distortion $D(Q,F)$. But we also have, by virtue of uniform integrability, that

$$\begin{aligned} D(Q,F) &= \int C(x - Q(x))\, dF(x) \\ &\le \liminf \int C(x - Q_n(x))\, dF(x) \\ &< \infty . \end{aligned}$$

Since this contradicts (3), it follows that there has to be a convergent subsequence of $\{Q_n\}$. In summary, except when $C(x)$ is a constant, the assumption in the above theorems that $\{Q_n\}$ converges (or has a convergent subsequence) is satisfied.

III. APPLICATIONS

The conditions hypothesized above may arise when an N-level scalar quantizer is to be designed for an unknown scalar distribution F, with $C(x) = |x|^r$, $r > 0$ (r^{th}-power distortion). One approach to this problem is to take n random samples from F and form the empirical distribution $\hat{F}_n(x)$ = (# of samples $\leq x)/n$. By the Glivenko-Cantelli Theorem [10] $\hat{F}_n(t) \to F(t)$ uniformly in t almost surely as $n \to \infty$. To apply the theorem in Section 2, we have to show that $|x-y|^r$ is absolutely integrable with respect to $\{\hat{F}_n\}$. By Kolmogorov's Strong Law of Large Numbers [7, p.239]

$$\lim \int_{|x|>a} |x-y|^r \, d\hat{F}_n(x) = \int_{|x|>a} |x-y|^r \, dF(x) \quad \text{almost surely if}$$

$\int |x|^r \, dF(x) < \infty$. Let $\varepsilon > 0$ be given. For any sample sequence for which the above limit holds we may choose a so that the right hand side is less than $\varepsilon/2$, and n_0, depending on the sample sequence, so that

$$\left| \int_{|x|>a} |x-y|^r \, d\hat{F}_n(x) - \int_{|x|>a} |x-y|^r \, dF(x) \right| < \varepsilon/2$$

whenever $n \geq n_0$. Then

$$\int_{|x|>a} |x-y|^r \, d\hat{F}_n(x) < \varepsilon \tag{4}$$

whenever $n \geq n_0$. It therefore follows that $|x-y|^r$ is uniformly integrable with respect to $\{\hat{F}_n(x)\}$ almost surely provided that $\int |x|^r \, dF(x) < \infty$.

It is relatively straightforward to show that an optimal quantizer exists for a distribution with compact support (such as an empirical distribution). Since a quantizer Q is specified by its output levels $(y_1, y_2, \ldots, y_N)$, the distortion $D(Q,F)$ may be considered a function of these levels. For continuous cost functions, this mapping is continuous. Without loss of generality we can restrict $y_1, y_2, \ldots, y_N$ to the convex hull of the support of F. From these considerations, it follows that there is some set of levels $(y_1, y_2, \ldots, y_N)$ which minimizes $D(Q,F)$, and this set describes an optimal N-level quantizer.

If Q_n is an optimal N-level quantizer for the empirical distribution $\hat{F}_n$, then as discussed in Section 2, there is a subsequence of $\{Q_n\}$ which converges to an optimal N-level quantizer for F. Thus we have managed to demonstrate, by sequential arguments, that optimal N-level quantizers exist for general scalar distributions, whenever (a) $C(x) = |x|^r$ and (b) $\int |x|^r \, dF(x) < \infty$. This result can be extended to continuous cost functions for which $C(x) =$

$O(|x|^r)$ for some positive r and $\int |x|^r \, dF(x) < \infty$. This type of argument has been used previously to show the existence of minimum mean-square error quantizers for the Laplace density [3].

Proceeding to k-dimensional vector quantizers, denote the scalar components of a vector x by $(x^1, x^2, \ldots, x^k)$. Similarly the marginal distributions of a probability distribution F on $\mathbb{R}^k$ will be denoted $F^1, F^2, \ldots, F^k$. Let $I = \mathop{\mathsf{X}}_{i=1}^{k} [-a,a]$ be a closed cell. Using an r^{th}-power distortion measure, we have

$$\int_{I^c} C(x-y)dF(x) = \int_{I^c} [\sum_{i=1}^{k} (x^i - y^i)^2]^{r/2} \, dF(x)$$

$$\leq k^{r/2} \sum_{i=1}^{k} \int_{|x^i|>a} |x^i - y^i|^r \, dF^i(x^i). \tag{5}$$

From (4) and (5) we see that multi-dimensional empirical distribution functions also result in almost surely uniformly integrable sequences, provided that the r^{th}-moments of the marginal distributions exist or, equivalently, that $\int ||x||^r \, dF(x) < \infty$. The succeeding developments in the scalar case generalize readily to several dimensions, including the Glivenko-Cantelli Theorem [2].

Another situation in which sequences of distributions arise is in the design of optimal vector quantizers using variational or fixed-point algorithms. Earlier we discussed an algorithm designed to circumvent the problem of convergence to locally optimal solutions. Linde, et al. [5] conjectured that this modified procedure converges to a global optimum, but were unable to prove it rigorously. Theorem 2.1 indicates that the modified algorithm is feasible if the uniform integrability criterion can be satisfied. Moreover, the theorems imply that if two distributions are "close" enough, then their respective optimal output levels are also close to each other. Thus it makes sense to use an optimal quantizer for one distribution as an approximation to the optimal quantizer for the other. The description given above is purposely vague on how close the distributions have to be, since this depends on the particular algorithm used.

Let F denote the distribution of the signal we wish to quantize. A specific choice of F_n recommended by Linde, et al. is the distribution of $W_n = (1-a_n)X + a_n Z$ where X has distribution F, Z has an arbitrary distribution G, and $\{a_n\}$ is a sequence decreasing monotonically to zero. We consider here the r^{th}-power cost function $C_0(t) = t^r$, and assume that

$$E\{||Z||^r\} < \infty \ , \ E\{||X||^r\} < \infty. \tag{6}$$

It follows from the above that $W_n \to X$ in r^{th}-mean, i.e.

$$\lim_{n\to\infty} E\{||W_n - X||^r\} = \lim_{n\to\infty} a_n^{\ r} E\{||X-Z||^r\} = 0.$$

This implies that $\{F_n\}$ converges weakly to F [4, p.452]. Using Minkowski's Inequality or the c_r-inequality it can be shown that

$$\lim E\{\|W_n-p\|^r\} = E\{\|X-p\|^r\} \tag{7}$$

for any vector p in $\mathbf{R}^k$. This is equivalent to saying that $C(x-p) = \|x-p\|^r$ is uniformly integrable with respect to $\{F_n\}$[4, p.138]. All of the conditions of Theorem 2.1 have been verified. Therefore, the sequence $\{Q_n\}$ of optimal quantizers for the F_n's must have a convergent subsequence, and the limit quantizer(s) is optimal for F. Note that except for (6), no other restrictions have been imposed on the distributions F and G. This allows the user considerable latitude in choosing G.

Another suggestion by Linde, et al. [5] is to use the distributions F and G as follows. Take random samples $X_1, X_2, \ldots$ and $Z_1, Z_2, \ldots$ from F and G, respectively, and form the empirical distribution $\hat{F}_n$ of the set of observations $\{(1-a_i)X_i + a_iZ_i : i=1,2,\ldots, n\}$. These may be thought of as independent samples of the random vector W_n. If $x_i = (x_i^1, x_i^2, \ldots, x_i^k)$ and $t = (t^1, t^2, \ldots, t^k)$ are vectors, we will use the notation $x_i \leq t$ as shorthand for the simultaneous set of inequalities $x_i^j \leq t^j$, $j=1,2,\ldots,k$. Then the empirical distribution function $\hat{F}_n$ may be written as

$$\hat{F}_n(t) = \frac{\sum_{i=1}^{n} \chi_t[(1-a_i)X_i + a_iZ_i]}{n} \tag{8}$$

where χ_t *is the indicator function* $\chi_t[x] = \begin{cases} 1 & x \leq t \\ 0 & \text{else.} \end{cases}$ *Observe that* $E\{\chi_t[(1-a_i)X + a_iZ]\} = P\{(1-a_i)X + a_iZ \leq t\} \to F(t)$ at all continuity points t of F. Applying the Strong Law of Large Numbers and the Toeplitz Lemma [4, p.89] to (8) yields

$$\lim_{n\to\infty} \hat{F}_n(t) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} P\{(1-a_i)X + a_iY \leq t\}}{n} = F(t) \quad \text{almost surely}$$

for all t in the continuity set of F. Hence $\hat{F}_n$ converges weakly to F, almost surely.

It remains to be shown that $C(x-p) = \|x-p\|^r$ is uniformly integrable with respect to $\{F_n\}$. This is equivalent to showing that [4, p.138]

$$\lim \int \|w-p\|^r \, d\hat{F}_n(w)$$

$$= \lim \frac{\sum_{i=1}^{n} \|(1-a_i)X_i + a_iZ_i - p\|^r}{n} \quad \text{almost surely}$$

$$= \int \|x-p\|^r \, dF(x). \tag{9}$$

Here we will assume that

$$E\{\|X\|^{2r}\},\ E\{\|Z\|^{2r}\} < \infty\ . \tag{10}$$

Then equation (7) holds. Applying the same reasoning as that used to derive (7), we have $\lim E\{\|(1-a_i)X + a_iZ - p\|^{2r}\} = E\{\|X-p\|^{2r}\}$.

This implies that $\sum_{i=1}^{\infty} \frac{E\{||(1-a_i)X + a_i Z-p||^r\}}{i^2} < \infty$ and, with (7), that

$$\sum_{i=1}^{\infty} \frac{Var\{||(1-a_i)X + a_i Z-p||^r\}}{i^2} < \infty \ . \qquad (11)$$

This is the condition for a Strong Law of Large Numbers to hold for (9).

Thus we have [7, p.238] $\lim \frac{\sum_{i=1}^{n} ||(1-a_i)X_i + a_i Z_i-p||^r}{n}$

$$= \lim \frac{\sum_{i=1}^{n} E||(1-a_i)X + a_i Z-p||^r}{n} = E\{||X-p||^r\}$$

almost surely.
The last line follows from the Toeplitz Lemma. This verifies (9) and therefore the uniform integrability requirement of Theorem 2.1 is satisfie We conclude that the sequence $\{Q_n\}$ of optimal quantizers for the $\hat{F}_n$'s mus have a convergent subsequence, and the limit quantizer is optimal for F. Observe that for this second modification to work, the somewhat stronger condition (10) was assumed.

IV. CONCLUSION

We have established conditions guaranteeing the convergence of a sequence of quantizers. These conditions were then used to establish the existence of optimal r-th power distortion quantizers. Also the convergence of the design techniques proposed by Linde, Buzo, and Gray was established.

APPENDIX

The following convergence theorem is proved.

Theorem A.1. Suppose that $\{F_n\}$ is a sequence of distribution functions on $\mathbb{R}^k$ converging weakly to a distribution function F, and that $g_n : \mathbb{R}^k \to \mathbb{R}$ is a sequence of continuous, real-valued functions converging to a (continuous) function g uniformly on compact cells. Suppose further that $\{g_n\}$ is uniformly integrable with respect to $\{F_n\}$. Then

$$\lim \int g_n dF_n = \int g \, dF.$$

Proof: Let $I = \overset{k}{\underset{i=1}{X}} [-a,a]$ be a cell such that F has no discontinuities on the boundary of I. Then

$$\left| \int_I |g_n| dF_n - \int_I |g| dF \right| \leq \int_I |g_n - g| dF_n + \left| \int_I |g| dF_n - \int_I |g| dF \right| .$$

The first term on the right hand side goes to zero by uniform convergence, and the second term goes to zero by the Helly-Bray theorem [10, p.83]. Therefore the left-hand side converges to zero, and it follows that

$$\int_I |g| dF \leq \lim \ \inf \int |g_n| dF_n < \infty \ .$$

Letting $a \to \infty$, we can invoke Fatou's Lemma [9, p.226] to show that the limit function g is integrable with respect to F. Now we have

$$\left|\int g_n dF_n - \int g dF\right| \leq \int_I |g_n - g| dF_n + \left|\int_I g_n dF_n - \int_I g dF\right|$$

$$+ \int_{I^c} |g_n| dF_n + \int_{I^c} |g| dF.$$

Letting n and a go to ∞, we get the terms on the right hand side to decrease to zero by virtue of uniform convergence, the Helly-Bray Theorem, uniform integrability, and integrability of g, respectively. This gives the desired result.

ACKNOWLEDGEMENT

This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-81-0047.

REFERENCES

[1] J.D. Bruce, On the Optimum Quantization of Stationary Signals, IEEE Int. Conv. Rec. Pt. I, pp. 118-124, 1964.

[2] L.P. Devroye, A Uniform Bound for the Deviation of Empirical Distribution Functions, J. Multivariate Anal., 7(1977), pp. 594-597.

[3] P. Fleischer, Sufficient Conditions for Achieving Minimum Distortion in a Quantizer, IEEE Int. Conv. Rec. Pt. I, pp. 104-111, 1964.

[4] R.G. Laha and V.K. Rohatgi, Probability Theory, John Wiley and Sons, New York, 1979.

[5] Y. Linde, A. Buzo and R.M. Gray, An Algorithm for Vector Quantizer Design, IEEE Trans. Comm., COM-28 (1980), pp. 84-95.

[6] S.P. Lloyd, Least Squares Quantization in PCM, Bell Telephone Laboratories Paper, Murray Hill, N.J., 1957.

[7] M. Loeve, Probability Theory, D. Van Nostrand Company, Inc., Princeton, 1963.

[8] J. Max, Quantizing for Minimum Distortion, IRE Trans. Inform. Theory, IT-6 (1960), pp. 7-12.

[9] H.L. Royden, Real Analysis, MacMillan Publishing Co., Inc., New York, 1968.

[10] H.G. Tucker, A Graduate Course in Probability, Academic Press, New York, 1967.

LMED

-8